

siRNA EXPRESSION SYSTEM

The present invention concerns recombinant siRNA vectors based on the regulatory regions of the U1 snRNA gene.

5 RNA interference is a process of sequence-specific post-transcriptional gene silencing highly conserved in evolution. The mediators of sequence-specific messenger RNA degradation are 21-23 nucleotide small interfering RNAs (siRNAs) generated by ribonuclease III cleavage from longer dsRNAs. Such a process, initially described in *C. elegans* (Fire *et al.*, 1998), and subsequently reported in insects, plants and fungi, has
10 been recently reproduced in mammalian cells through the use of short double-stranded RNAs (21-23 base pairs; Elbashir *et al.*, 2001). Because the efficacy of such molecules is very high, siRNA duplexes may be useful for targeted gene inactivation in human and nonhuman cells and may lead
15 to the development of therapeutics against viral or genetic diseases.

Recently, several groups have reported systems designed to introduce siRNAs in mammalian cells through transfection of oligonucleotides (Elbashir *et al.*, 2001; McManus and Sharp, 2002). Since the use of artificial siRNAs has the major drawback of requiring periodic
20 administrations, there is an obvious need to produce plasmids encoding siRNAs in order to obtain a long-term effect.

So far, all siRNA vectors rely on pol III-dependent promoters, such as the U6 snRNA (Lee *et al.* 2002; Paul *et al.*, 2002) or the H1-RNA (Brummelkamp *et al.*, 2002) genes and more recently a tRNA expressing
25 cassette (Kawasaki and Taira, 2003), patent application WO03/0006477.

The authors of the present invention have developed a novel plasmid vector based on the pol II-dependent regulatory regions of the U1 snRNA gene. The U1 promoter is active in all cell types and induces the accumulation of high levels of transcripts. In addition, the presence of a 3'
30 element responsible for the correct 3' end formation of U1 snRNA allows efficient and precise 3' end formation of the transcript.

The constructs of the present invention have the following advantages: i) they have no or very little sequence requirement for 5' and 3' end.

formation; ii) the pre-siRNA is rapidly exported to the cytoplasm where it is efficiently converted to the mature form; iii) cloning can be easily performed by means of ds-oligonucleotides, thus avoiding the PCR amplification step which very often produces aberrant products; iv) the U1

5 snRNA gene has a strong pol II promoter which does not undergo silencing in stable cellular clones. Moreover, the authors have identified specific elements that allow the selection of a one single siRNA strand as a response mediator to interference.

Hairpin sequences against the lamin A/C mRNA were inserted between

10 these regulative regions and produced an in vivo efficient accumulation of double-stranded siRNAs of the correct size. Western blot analysis of proteins extracted from HeLa cells transfected with the siRNA expressing plasmids have shown that these vectors, which are the objects of the invention, are very efficient in suppressing the expression of the lamin A/C

15 protein. The sequences that determine an asymmetric release of one of the two double-stranded siRNA have also been identified.

Hence, the object of the present invention is a recombinant vector for the correct, stable and effective expression in mammalian cells of a siRNA or of a miRNA, comprised from 5' to 3':

20 a) a polymerase II RNA dependent promoter sequence derived from the U1 snRNA gene;

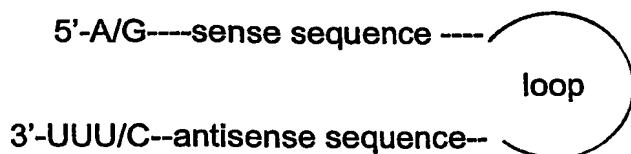
b) suitable restriction sites for cloning the sequence that transcribes a pre-siRNA or a pre-miRNA;

c) a sequence transcribing the pre-siRNA comprising in position +1 an A or a G residue; a sequence from 21 to 23 nucleotides corresponding to a

25 sense region of the mRNA transcribed by the gene to be silenced that constitutes the first segment of the stem of the pre-siRNA; a sequence selected from a pre-miRNA sequence that consists of the loop region of the pre-siRNA; a sequence from 21 to 23 nucleotides corresponding to the

30 antisense region of the mRNA transcribed by the gene to be silenced that constitutes the second segment of the stem of the pre-siRNA; two final residues UU protruding in such a way that the following structure is obtained:

3

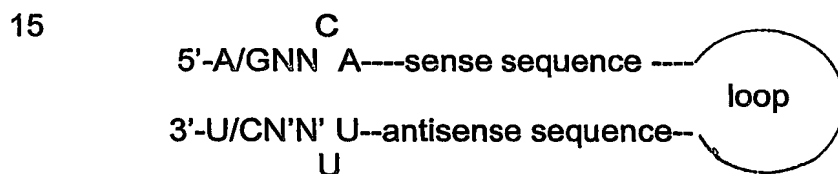


5 or alternatively a sequence transcribing a pre-miRNA;

d) termination sequences derived from the 3' end sequence of the gene for U1 snRNA which are necessary and sufficient for the correct formation of the 3' end of the pre-siRNA or of the pre-miRNA respectively.

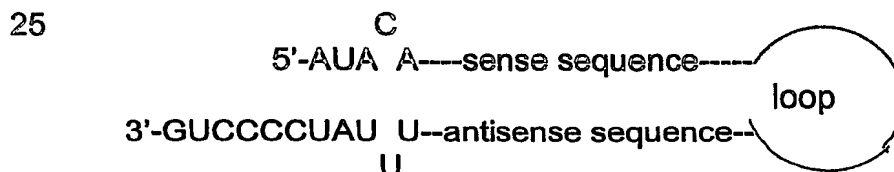
10 Preferably, the cloning site for the 5' of the sequence transcribing the pre-siRNA is Bgl II.

Preferably, the sequence transcribing the pre-siRNA further comprises at termini 5' and 3' such sequences that the transcribed pre-siRNA has the following structure:



where N is A, U, G or C and N' is its complementary nucleotide.

More preferably, the sequence transcribing the pre-siRNA further comprises at the 5' and 3' termini sequences such that the transcribed pre-siRNA has the following structure:



The vectors of the invention are advantageously utilized though they produce lower levels of siRNAs than pol III based vectors. As a matter of fact these levels result to be sufficient to mediate an RNA interference response, with no side effects due to unspecific targeting caused by high

transcripts levels. Then therapeutical applications are safer when the same response is obtained with such lower levels of siRNA.

The U1 snRNA gene promoter consists of a Distal Sequence Element (DSE) acting as enhancer of transcription and a Proximal Sequence Element (PSE) that positions transcription initiation. The distance between the PSE and the +1 nucleotide is strictly conserved while its primary sequence is not. This promoter organization is advantageous to modify and convert the same into an inducible promoter.

In the last few years it has been estimated that as many as 1 % of human genes encode a group of 20-25 nucleotide long non-coding microRNAs (miRNAs) that play a vital role in the regulation of gene expression in plant and animal species (Bartel, 2004). Their structure and biosynthesis is almost identical to that of siRNAs even though they are transcribed as longer precursors. The vector of the invention is advantageously use also to produce microRNAs.

siRNA expression systems applied to human gene therapy

The use of vectors for the in vivo expression of siRNAs finds very powerful applications as a therapeutic tool to generate RNAi response against selected disease models.

RNAi approaches can be used to knocking down gene expression in order to:

- 1) down-regulate deleterious gene expression
- 2) re-activate cellular defenses, and
- 3) inhibit viral gene expression.

This would enable treatments against many acute and persistent diseases including cancer and other genetic dysfunctions, as well as severe infections.

The effectiveness of the constructs produced against the selected target RNAs can be analyzed:

- 1) in transfected cells or in established cell lines.
- 2) in primary cells established from patients and/or in animal models of the specific genetic disease.

Animal models will contribute to the validation of siRNA vectors regarding optimized specificity, potency, stability, and delivery before starting clinical tests in human.

The invention is described below with examples given in reference to the following figures:

Figure 1. Schematic representation of the psiUX vector; sites present in the polylinker downstream from the U1 snRNA promoter are indicated. The sequence of the U1 promoter region extending from position -393 to position -6 with respect to the initiation site is shown below.

Figure 2. Panel A): sequence of different psiUx derivatives and of related oligos. 5'end and 3'end of the transcripts are indicated. Sense and antisense sequences are deduced from the lamin A/C mRNA and are represented by convergent arrows. The mutations introduced in the psiUc_{mut}-lam construct are indicated above the sequence. The 3' terminator sequence of the U1 gene is indicated as "3' box". Internal loops and variant sequences present in the psiUb-lam and psiUd-lam constructs are shown in grey. Panel B): Predicted structures of the four tested anti-lamin primary transcripts. The arrows indicate the presumptive sites of processing by the Dicer enzyme. siRNA sequences are shown in bold and italics. Underlined nucleotides identify the sequences derived from the 5' and 3' regions of U1 snRNA. The asterisk represents the monomethyl cap.

Figure 3. Analysis of the expression and activity of the siRNAs transcribed by the psiUx-lam constructs. HeLa cells were transfected with 6 μ g of the different psiU-lam constructs (lanes psiU) or with 6 μ g of U6-lam (lane U6). 2 μ g of a control U7 construct were co-transfected in all cases. After 48 hours, total RNA was extracted and 15 μ g were analysed by Northern blot on a 10% polyacrylamide-urea gel. Panels A and A'): hybridization with *a* and *a_{mut}* probes that recognize the antisense strand (left-pointing arrow). Panel A''): hybridisation with a probe specific for U7 snRNA. The brackets indicate the position of the precursor species. The migration of a pBR322/MspI molecular size marker is shown on the left. Lane NT contains RNA extracted from untransfected cells. Panel B): hybridization with an *a* probe that recognizes the sense strand (right-pointing arrow).

Panel c): 20 μ g of total cellular proteins extracted 70 hours after transfection from the same cells as in the previous experiment were analysed by Western blot analysis with anti-lamin monoclonal antibodies. The arrows point to the two isoforms of lamin (A and C). Below the panel, the Ponceau staining of the filter is shown.

Figure 4. Analysis of the expression of psiUa-lam, psiUb-lam and U6-lam transcripts in *Xenopus laevis* oocytes. 2.76 ng of plasmid DNA were microinjected into the nucleus of *Xenopus laevis* oocytes. After 12 hours of incubation, RNA was extracted from nuclei (N) and cytoplasms (C) and analysed by Northern blot on a 10% polyacrylamide-urea gel. The arrows indicate the primary transcripts. The brackets indicate the products of cleavage inside the loop (see schematic representation on the side).

Materials and methods

Construction of psiUx vector. The U1 snRNA gene-based vector was derived from plasmid pHU1-ID, containing the entire human gene (De Angelis *et al.*, 2002); this plasmid carries the 600 bp *Bam*HI fragment containing the transcriptional unit of the human U1 snRNA gene inserted in the *Bam*HI site of the pSP65 vector (Promega) in the opposite direction of the SP6 promoter. Plasmid psiUx was derived from the latter by double digestion with *Bgl*II and *Nhe*I and religation in the presence of a polylinker containing the 5'-*Bgl*II, *Kpn*I, *Xho*I, *Nhe*I, *Bam*HI and *Nhe*I-3' sites. The *Bgl*II site maps in the U1 snRNA gene, at position -6 with respect to the transcription initiation site, while the *Nhe*I site is in the vector, 300 nucleotides upstream from the SP6 promoter. The linker was made by annealing the two oligonucleotides:

linkup: 5'-GATCTGGTACCCTCGAGGCTAGCGGATCCG-3'

linkdn: 5'-CTAGCGGATCCGCTAGCCTCGAGGGTACCA-3'.

Construction of psiU-derivatives expressing siRNAs for the lamin A/C protein.

The selected target sequence on the lamin A/C was derived from Sui *et al.* (2002) and covers nucleotides 1602-1622 of the X03444 of the NCBI Database. The following oligos:

a-lamUP

5 5'GATCTCATACAGGGCAATTGGCAGATCAAGCGTTGTGAAGCCACAG
ATGAACGCTTGATCTGCCAATTGCCCTTTATCCCCTGACTTTCTGGAG
TTTCAAAGTAGAC3'

and a-lamDN

5'TCGAGTCTACTTTTGAAACTCCAGAAAGTCAGGGGATAAAGGGCAA
10 TTGGCAGATCAAGCGTTCATCTGTGGCTTCACAACGCTTGATCTGCC
AATTGCCCTGTATGA3'

were annealed and inserted in the *Bgl*II and *Xho*I sites of psiUx, giving rise to plasmid psiUa-lam. Plasmid psiUb-lam and psiUc-lam were obtained by cloning in the *Bgl*II and *Xho*I sites of psiUx the following oligos:

15 psiUb-lam:

b-lamUP

5'GATCTCATACAGGGCAATTGGCAGATCAAGCGTTTGTGTAGCGCTT
GATCTGCCAATTGCCCTTTATCCCCTGACTTTCTGGAGTTTCAAAGT
AGAC3'

20 and b-lamDN

5'TCGAGTCTACTTTTGAAACTCCAGAAAGTCAGGGGATAAAGGGCAA
TTGGCAGATCAAGCGCTACACAAACGCTTGATCTGCCAATTGCCCTG
TATGA3'

psiUc-lam:

25 *c-lamUP*

5'GATCTCGGGCAATTGGCAGATCAAGCGTTTGTGTAGCGCTTGATCT
GCCAATTGCCCTTACTTTCTGGAGTTTCAAAGTAGAC3'

and c-lamDN

5'CTGAGTCTACTTTTGAAACTCCAGAAAGTAAGGGCAATTGGCAGAT
30 CAAGCGCTACACAAACGCTTGATCTGCCAATTGCCCCGA3'

psiUd-lam:

d-lamUP

5'GATCTCGGGCAATTGGCAGATCAAGCGTTTGACTTCGCATGAATGA

GTTTCATTCATGAAGCGAAACGCTTGATCTGCCAATTGCCCTTACTTTC
TGGAGTTTCAAAAGTAGAG3'

and d-lamDN

5'CTAGCTCTACTTTTGAAACTCCAGAAAGTAAGGGCAATTGGCAGAT
5 CAAGCGTTTCGCTTCATGAATGAACTCATTTCATGCGAAGTCAAACGCT
TGATCTGCCAATTGCCCCGA3'.

Plasmid psiUc_{mut}-lam was obtained by cloning oligos:

cmut-lamUP

5'GATCTCGGGCAATTGcgAGATCAAGCGTTTGTGTAGCGCTTGATCTc
10 gCAATTGCCCTTACTTTCTGGAGTTTCAAAAGTAGAC3'

and cmut-lamDN

5'CTGAGTCTACTTTTGAAACTCCAGAAAGTAAGGGCAATTGcgAGATC
AAGCGCTACACAAACGCTTGATCTcgCAATTGCCCCGA3'

(lower case letters indicate nucleotides mutated with respect to the lamin
15 sequence).

Cell culture and transfection. Subconfluent HeLa cells were transfected
in 60 mm plates by using Lipofectamine 2000 (Life Technologies, Gibco
BRL) according to the manufacturer's instructions. 6 µg of psiUx plasmid
derivatives were transfected together with 2 µg of plasmid U7-3'
20 (DeAngelis *et al.*, 2002) as an internal control of transfection.

***Xenopus laevis* oocytes microinjections.** 9.2 nl of plasmid DNA (300
ng/µl) were injected in the nucleus of stage IV *Xenopus laevis* oocytes
according to Caffarelli *et al.* (1987). After 12 hours of incubation at 19 °C,
nuclei and cytoplasms were manually dissected and RNA was extracted
25 as described (Caffarelli *et al.*, 1987).

Northern blotting. Total RNA isolation from transiently transfected HeLa
cells was done using the Ultraspec RNA isolation system (Biotech
Laboratories, Houston) according to the manufacturer's instructions. To
detect siRNAs, 15 µg of total RNA were electrophoresed in a 10%
30 polyacrylamide-8 M urea gel and transferred by electroblotting onto
Hybond-N⁺ membrane (Amersham Pharmacia Biotech.). The hybridisation
was carried out at 37 °C in 5X SSPE, 5X Denhardt's solution, 0.5% SDS,

25 ug/ml salmon sperm DNA (Invitrogen). Washes were performed at 37 °C in 6X SSPE and 2X SSPE and 0.2X SSPE. Probes used were terminally ³²P-radiolabelled DNA oligos:

probe a: 5'-GGCAATTGGCAGATCAAGCG-3';

5 probe a-mut: 5'-GGCAATTGcgAGATCAAGCG-3';

α probe: 5'-CGCTTGATCTGCCAATTGCC-3'.

U7-3' transcript was detected with probe U7a (DeAngelis *et al.*, 2002).

Immunoblotting. Protein extracts (20 μg) were separated on 10% polyacrylamide-SDS gels and transferred to nitrocellulose (ProTran, Schleicher and Schuell). The membranes were blocked with 3% skimmed milk in TBS. Mouse monoclonal anti-lamin A/C antibody (sc-7292, Santa Cruz Biotechnology) diluted 1:200 in TBS/3% skimmed milk was used as primary antibody. The immunostaining was carried out using ECL Western blotting detection system (Amersham UK).

15 **Results**

The system takes advantage of the characteristics of the human U1 snRNA gene and of its promoter and terminator regions (Hernandez, 1985, Hernandez and Weiner, 1986). The U1 promoter regulates transcription by RNA polymerase II, is ubiquitously active, and ensures high levels of expression. The primary transcript has a monomethylated cap and the RNA is efficiently exported to the cytoplasm. This is extremely important for efficient processing of the pre-siRNA, since the Dicer enzyme has been shown to localize in the cytoplasm (Billy *et al.*, 2001). In addition, the correct 3' end formation of U1 snRNA is directed by a box element (GTTTCAAAAGTAGAC-3'box) located 10 nucleotides downstream from the U1 snRNA coding region, which works only in association with the specific U1 promoter sequence (Hernandez and Weiner 1986; de Vegvar *et al.*, 1986). A similar sequence has been found to direct correct 3' end formation also of the U2 snRNA (Hernandez, 1985). In this respect it has been suggested that these snRNAs must be transcribed by a specialized transcription machinery that differs from that synthesizing mRNAs. Recent work by Medlin *et al.* (2003) has shown that termination does not occur

properly if the CTD of pol II has been deleted, indicating that factors required for 3' end formation are recruited very early during transcription. The region containing the U1 snRNA gene promoter extending from the BamHI site, at position -400, to the BglII site, at position -6 with respect to the initiation site, was cloned into the BamHI/NheI sites of the pSP65 plasmid through the use of a synthetic polylinker (Figure 1A). The resulting construct (psiUx) has a very strong promoter but lacks the transcription initiation site and the terminator. According to our design, these sequences should be provided by the inserted synthetic double stranded fragment, together with the siRNA target sequences. The initiation sequence is 5'-GATCTC'A-3', where the last residue corresponds to the +1 nucleotide of the U1 snRNA (a G is also accepted in this position). The terminator element is 5'-CCCCTG'ACTTTCTGGAGTTTCAAAGTAGAC-3', where the underlined sequence is the so-called 3' box, located 10 nucleotides downstream from the 3' end of the transcript. The CCCCTG sequence corresponds to the last 6 transcribed nucleotides of the U1 snRNA which have been shown to contribute to an efficient and site-specific 3' end formation (Hernandez, 1985). Cloning of the siRNA precursor sequence into psiUx can be performed very easily by inserting an amplified fragment or annealed synthetic oligonucleotides with ends compatible with the selected sites of the plasmid. While a 5' BglII terminus is obligatory since it is required to restore the initiation site, any of the sites contained in the polylinker (KpnI, XhoI, NheI and BamHI) can be utilized at the 3' end (Fig. 1).

For a target sequence to test the effectiveness of our vector we selected a site in the lamin A/C mRNA demonstrated to be vulnerable to siRNAs (Sui *et al.*, 2002). A hairpin (21 nucleotides sense and antisense sequences derived from the lamin A/C mRNA) was cloned in different contexts in order to identify the most appropriate for efficient siRNA expression (psiU-lam constructs). The resulting constructs differ not only in the type of loop sequence inserted but also in the 5' and 3' ends of the transcribed region. The sequence of the different inserted fragments is indicated in Fig.2A, and the structure of the primary transcripts is schematically represented in

panel B. The loops utilized in psiUa-lam, psiUb-lam and psiUc-lam are deduced from micro pre-RNAs (Zeng *et al.*, 2002 and Castanotto *et al.*, 2002), while that present in psiUd-lam is deduced from the canonical substrate of the yeast Rnt1p endonuclease, a RNaseIII-like enzyme (Chanfreau *et al.*, 1998). In addition, constructs psiUa-lam and psiUb-lam contain a terminal region comprising a stem of 3 nucleotides that correspond to the conserved nucleotides at the 5' and 3' ends of the U1 transcript. To prevent Dicer from cleaving in this region, two non-matched bases were inserted. These extensions are absent in psiUc-lam and psiUd-lam. In these constructs, to match the lamin target sequence, the 5' end of the transcript included a G, and the conserved 3' region was converted from CCCCTG to CCCTT.

A control construct was derived with two mismatches of two nucleotides in the central part of the lamin mRNA pairing region from psiUc-lam (psiUc_{mut}-lam). This siRNA produced by this construct should be unable to mediate interference response.

To compare the activity of these vector systems with others previously utilized, an anti-lamin A/C hairpin sequence in the U6 vector (plasmid U6-lam) was cloned as described by Sui *et al.* (2002). The different constructs were tested for expression and activity by transfection into HeLa cells. As an internal control, a modified U7 snRNA gene was co-transfected (DeAngelis *et al.*, 2002) At 48 hours, RNA was extracted and analysed by Northern blot.

Panels A and A' in Figure 3 show hybridizations with a specific probe for the sense antisense strand (probe a). All U1-derived constructs produce an accumulation of the antisense strand ranging between 21 and 23 nucleotides. Since these RNAs accumulate in vivo over prolonged times, it can be argued that they are present in stable complexes. Similar findings have been interpreted in other cases as diagnostic for the association with an interference-competent complex. The same type of molecules, in the range of 21-23 nucleotides, accumulates over time in the case of the U6-derived constructs (lane U6). The comparison between the levels of U6 and U1-derived siRNAs, after normalization to the hybridization signal of

the co-transfected U7 snRNA (panel A''), indicates that the transcriptional level of psiU-lam constructs is slightly lower than U6-lam. Nevertheless, as shown in Fig. 3C, they mediate the same level of RNA interference response. The siRNA derived from psiUC_{mut}-lam are only
5 visible with the *a_{mut}* probe, which has perfect complementarity to the mutation (Fig. 3A').

The hybridizations performed with the antisense probe (*probe α* , panel B) reveal an interesting feature: the psiUa-lam and psiUb-lam constructs show no accumulation of the sense strand (the upper strand in Fig. 2B)
10 also when the gel is exposed for lengthy periods. In contrast, all the other constructs showed the production of the sense strand, even if the levels were lower than those of the antisense strands (precursor compared with mature species hybridization signals in panels A and B). These results indicate that the terminal regions of the psiUa-lam and psiUb-lam
15 constructs (Fig. 2B), but not the internal loop, are the element that confers asymmetric selection to the strand included in the interference complex.

Expression analysis of the psiU (Fig. 3A) and other independently produced constructs indicate that it is possible to eliminate most of the conserved nucleotides at the 3' end of the U1 snRNA and still obtain
20 efficient termination and processing. Hence, the cloning into the psiUx vector has no sequence constraints, other than the presence of an A or a G at position +1.

An interesting difference between the polIII and polII siRNA vectors is that in the case of U1-driven transcripts only low amounts of precursor
25 (indicated as pre-siRNA) can be detected. On the contrary, this species is much more abundant in the case of the U6 vector (Fig. 3A).

A possible explanation for this difference is that the U6-driven transcripts are not efficiently exported to the cytoplasm. To verify this hypothesis, the U6-lam and psiUa-lam and psiUb-lam plasmids were microinjected into the
30 nucleus of *X. laevis* oocytes, and after 12 hours of incubation RNA was extracted from the nuclear and cytoplasmic compartments. Fig.4 indicates that a large proportion of the U6-driven transcripts are retained in the nucleus, while no trace of the U1-driven products is found in there. In both

cases, only RNA transcripts of the size expected for a precursor molecule are found (pre-siRNA) together with shorter RNA species of the size expected if cleavage had occurred at any position inside the loop region. In the *Xenopus* system, only tiny amounts of 21-23 nucleotide long species can be visualized after long overexposures, indicating that a Dicer-like activity, if present, is in very low amounts. These data indicate that the pre-siRNAs produced under the U1 promoter are efficiently exported to the cytoplasm.

The same cells analysed for siRNA production were also tested for RNAi activity. 20 micrograms of total protein extracts from cells transiently transfected with U1 and U6-derived constructs were analysed by Western blot analysis with anti lamin A/C antibodies. As shown in Fig.3B, all siRNA vectors produce good levels of interference if we consider that the efficiency of transfection is in the order of 80-85% (not shown). The level of lamin depletion is similar among all the U1 constructs and to those of the U6-derived expression cassette.

The specificity of the interference response was tested by transfecting into HeLa cells a construct containing two mismatches in the central part of the 21 nucleotide long pairing region (psiUc_{mut}-lam). While the accumulation levels of the siRNAs from this construct are similar to those obtained with the parental construct (Fig.3A') they do not mediate interference as indicated by the levels of lamin accumulation (fig.3B, lane mut) that are similar to the control (lane NT).

Altogether, these data indicate that the U1-based vectors have several interesting features compared with other siRNA vectors: i) they have very little sequence requirement at the 5' and 3' termini of the transcripts; ii) they accept U sequences in the transcribed region, unlike what occurs in pol III vectors; iii) cloning is very easy and allows the selection among different cloning sites; iv) primary transcripts are efficiently exported to the cytoplasm and converted to the mature form. In addition, specific sequences can be inserted at the 5' and 3' termini, thus allowing the selection of a single siRNA strand that must be incorporated into the interference complex. This is an important feature for a siRNA vector, in

that it eliminates the accumulation of the sense strand that could mediate undesired targeting.

Inducible U1-based expression vector

5 The U1 snRNA gene promoter consists of a Distal Sequence Element (DSE) acting as enhancer of transcription and a Proximal Sequence Element (PSE) that positions transcription initiation. The distance between the PSE and the +1 nucleotide is strictly conserved while its primary sequence is not. The promoter organization is advantageous to modify and convert the same into an inducible promoter.

10 One example is represented by the Cre-loxP strategy (Hoess, R.H., A. Wierzbicki, and K. Abremski, 1986; Sauer, B. and N. Henderson, 1988). A long stuffer DNA (containing the information for a selectable marker, i.e. puromycin, neomycin or others), flanked by 2 loxP sites of bacteriophage P1 can be inserted between the PSE and the +1 position. Such insertion
15 prevents transcription of the siRNA. Upon addition of the Cre recombinase the two LoxP sites recombine with the resulting deletion of the stuffer DNA and of one loxP site. Provided that the distance between the PSE and the +1 nucleotide is maintained, the presence of one loxP site does not interfere with promoter activity and transcription of the pre-siRNA occurs.

20 Expression of microRNAs.

miRNA function has been correlated with many complex cellular circuits such as control of cell proliferation, cell death, and fat metabolism in flies, neuronal patterning in nematodes, modulation of hematopoietic lineage differentiation in mammals, and control of leaf and flower development in
25 plants. Recently, a role for microRNAs in oncogenesis has also been proposed. Mainly from studies in worm and flies it has been shown that microRNAs are processed and functionalized in the cell via the RNAi pathway, and that through base pairing to the 3' untranslated region of a mRNA it causes a significant decline in its expression.

30 The U1-based vector described in this work is utilized also for the production of high levels of miRNAs. In order to optimize miRNA expression two different strategies are followed: 1) the pri-miRNA sequence (this sequence includes a region of approximately 80-110

nucleotides with a well defined secondary structure and can be deduced from genomic data (Griffiths-Jones et al. 2003) is cloned between the promoter and the termination region of the psiUx vector; 2) the miRNA sequence is cloned in psiU as an antisense siRNA sequence.

5

Altogether, these data indicate that the U1-based vectors have several interesting features compared with other siRNA vectors: i) they have very little sequence requirement at the 5' and 3' termini of the transcripts; ii) they accept U sequences in the transcribed region, unlike what occurs in pol III vectors; iii) cloning is very easy and allows the selection among different cloning sites; iv) primary transcripts are efficiently exported to the cytoplasm and converted to the mature form; v) specific sequences can be added at the 5' and 3' termini, that allow the selection of only one of the two siRNA strands to be incorporated into the interference complex. This is an important feature for a siRNA vector, in that it eliminates the accumulation of the sense strand that could mediate undesired targeting. This is a very crucial aspect that ensures high levels of safety if this expression system has to be used in human clinical protocols.

20

References

Bartel D. P. *Cell*. 2004, 116(2):281-97.

Billy, E., et al. *Proc Natl Acad Sci U S A* 25,14428-33 (2001).

Brummelkamp, T. R., Bernards, R. & Agami, R. *Science* 296, 550-553 (2002).

25

Caffarelli, E., et al. *EMBO J.* 6, 3493-8 (1987).

Castanotto, D., Li, H. & Rossi, J.J. *RNA*. 8,1454-1460 (2002).

Chanfreau, G., Legrain, P. & Jacquier, A. *J. Mol. Biol.* 284, 975-988 (1998).

De Angelis, F.G., et al. *Proc Natl Acad Sci U S A*. 99, 9456-9461 (2002).

30

de Vegvar, H.E., Lund, E. & Dahlberg, J.E. *Cell* 47,259-266 (1986).

Elbashir, S.M., et al. *Nature* 411, 494-498 (2001).

Fire, A., et al. *Nature* 391,806-811 (1998).

Griffiths-Jones S, et al. *Nucleic Acids Res.* 2003 31(1):439-41.

- Hernandez, N. *EMBO J.* **7**,1827- 1837 (1985).
- Hernandez, N. & Weiner, A.M. *Cell* **47**,249-258 (1986).
- Hoess, R.H., A. Wierzbicki, and K. Abremski,. *Nucleic Acids Res*, 1986. **14**(5): p. 2287-300.
- 5 Kawasaki, H. & Taira, K. *Nucleic Acids Res.* **31**, 700-707 (2003).
- Lee, N.S., et al. *Nat. Biotechnol.* **20**,500-505 (2002).
- McManus, M.T. & Sharp, P.A. *Nat Rev Genet.* **10**,737-747 (2002).
- Medlin, J.E., et al. *EMBO J.* **22**,925-934 (2003).
- Paul, C.P., et al. *Nat. Biotechnol.* **20**, 505-508 (2002).
- 10 Sauer, B. and N. Henderson. *Proc Natl Acad Sci U S A*, 1988. **85**(14): p. 5166-70
- Sui, G. et al. *Proc Natl Acad Sci U S A* **99**, 5515-5520 (2002).
- Zeng, Y., Wagner, E.J. & Cullen, B.R. *Mol Cell.* **9**, 1327-1333 (2002).